Implementation of multipath network virtualization scheme with SDN and NFV

QingtianWang, Junli Xue, Guochu Shou, Yaqiong Liu, Yihong Hu, Zhigang Guo
Beijing Key Laboratory of Network System Architecture and Convergence
School of Information and Communication Engineering
Beijing University of Posts and Telecommunications, Beijing 100876, China
{qtwang, junlixue, gcshou, liuyaqiong, yhhu, gzgang}@bupt.edu.cn

***Abstract*—Multipath algorithms except Equal-Cost Multi-Path(ECMP) which has been widely used in networks are difficult to apply, because multipath provisioning is more complex at cross layers and multipath routing need to get all nodes' information. To address the dilemma, this paper proposes a multipath network virtualization implementation scheme with Software Defined Networking (SDN) and Network Function Virtualization (NFV). In this scheme, SDN schedules network resources in a global view for selecting multiple paths and computing weight of each path, and NFV provides computing and storage resources to split flow, add tag, recover flow, to name a few. This paper also proposes a multipath algorithm for elephant flow with network virtualization. Besides, we build an experimental platform based on OPNFV and SDN, and conduct experiments under this experimental platform. The results show that our proposed algorithm applied on multipath network virtualization experimental platform has superior performance than ECMP applied in networks without virtualization.**

***Keywords*—multipath; virtualization; NFV; SDN; elephant flow**

I. INTRODUCTION

The new online applications, such as social networking (e.g.,Twitter,Weibo), high definition video streaming (e.g.,YouTube) and serious game propose a new set of demands on throughput, bandwidth, Quality of Service(QoS) and end-to-end delay. In present Internet, most of communications are established over a single path, which is easy to cause network constraints (e.g., load unbalancing). Multiple paths no longer aim only at solving problems from a single path but also focus on ensuring high quality network services and guaranteeing QoE (Quality of Experience) [1][2]. Multipath configurations can be established in several different ways, [3] discusses five multipath cases in different networks, such as multipath routing at source node with multiple interfaces and multipath routing over wireless mesh network or mobile ad hoc.

Equal-Cost Multi-Path (ECMP) is simple to deploy in multipath routing, so it is widely used in present network and it spreads the flow over multiple equal cost paths using hash functions. Most of routing protocols, such as Open Shortest Path First (OSPF) [4] and Intermediate System to Intermediate System (IS-IS) [5], support ECMP. Besides, some router implementations also allow ECMP usage with Routing Information Protocol (RIP). However, some algorithms, such as FLARE [6], LBPF [7], LDM [8], which also have good performances are not widely used. The main reasons are that multipath provisioning is more complex at higher layers than at the lower layers, and network topological information need to be disseminated to all nodes in multipath routing [3].

Network virtualization can abstract physical layer resource to the virtual layer. Furthermore, the global view of network resources will be formed in virtual layer. The control layer manages network resources in an effective way, thereby enhancing the network resource utilization and avoiding complex signaling mechanisms at cross layers. In [9]S.He et al. proposed a viable way to establish multipath access in the FiWi network through the flexible use of virtual networks with network virtualization.

Software defined networking (SDN) is a network paradigm that separates the control and data plane and consolidates the control functions into controller. In SDN paradigm, all control functions are concentrated in the control plane, and controller manages network through the global view of network. Recently, there are lots of researches on multipath with SDN, such as [20] [21] [22]. SDN has the ability to simplify network design, so it is easy to deploy the new network mechanism. It has been widely adopted as a means to enable network virtualization [10]. However, network devices could not provide lots of computing and storage resources, so the emergence of Network Function Virtualization (NFV) and SDN forms a supplement.

NFV aims to address the dilemma of traditional hardware-based network devices. Network functions are softwarized and can be deployed dynamically on

978-1-5386-3531-5/17/$31.00 ©2017 IEEE

network to meet changing traffic and service usage levels. Therefore, these solve the problems that traditional network devices which are difficult to manage and change for traditional network devices [11]. Furthermore, the expansion of the network becomes easier to save network energy consumption. European Tele-communications Standard Institute (ETSI) has their own standard of NFV [12].Communications vendors, such as Huawei, Ericsson also claimed that their devices support NFV[13].The emergence of NFV provides a new direction for communication network. New services can be deployed rapidly in the network by using NFV. Thi-Thuy-Lien Nguyen develops a multipath routing solution for minimizing the maximum link utilization in NFV-based systems, as well as the efficient utilization of network resources [14]. The load balancing problem is solved by using multipath routing in NFV to optimize network performance [15][16]. As a new type of network technology, SDN and NFV provide the support for the realization of network virtualization.

In this paper, we use industrial servers to implement the network function based on the concept of NFV as well as deploying the SDN controller, NFV orchestrator and OpenFlow switches in SDN. This way is flexible to implement multipath network virtualization.

The contributions of our paper are listed as follows:

A. We propose the multipath network virtualization implementation scheme with SDN and NFV.

B. A Delay Minimization Multi-Path algorithm with network virtualization (DMMP) for elephant flow is proposed and embedded in our scheme.

C. We conduct the multipath network virtualization experimental platform. The results demonstrate that our proposed algorithm applied on multipath network virtualization experimental platform has superior performance than ECMP applied in traditional network.

The rest of the paper is organized as follows: Section II describes multipath network virtualization implementation scheme and some functional components. Section III presents DMMP algorithm for elephant flow. In section IV we conduct several experiments under our experimental platform and the results imply our scheme has superior performance. We conclude the paper and conceive the future work in section V.

II. MULTIPATH NETWORK VIRTUALIZATION IMPLEMENTATION SCHEME

In this section, we present the multipath network virtualization implementation scheme in detail, as shown in Fig. 1. The scheme is divided into control plane and data plane, and multipath is implemented in data plane.

In the control plane, SDN controller coordinates NFV orchestrator to manage the network, computing and storage resources, through network resource management component. All resources are provided by nodes of the data plane. The algorithm embedded in path selection and weight computation component for selecting multiple paths and computing weight of each path depending on the requirements of traffic and dynamic network state. After that, the tag table component translates the information of multiple paths to appropriate OpenFlow tables which include the tag to distinguish traffic units and sends the OpenFlow tables to the devices which support OpenFlow protocol to forward the traffic. The action set of OpenFlow tables also has Set-Queue action to provide basic Quality-of-Service (QoS) support.

In the data plane, traffic transmits through several functional components which are deployed by virtualization network functions (VNFs) and multiple paths are unifiedly scheduled by the control plane. In order to avoid the impacts of the failure of devices or links, the state and information of links and devices will update periodically. Data plane consists of four components, in which flow splitting, mapping and forwarding are deployed in the source node, and flow recovery is implemented in the destination node. Meanwhile, the computing and storage resources of all nodes are abstracted by virtualization and then the Virtualized Infrastructure Manager (VIM) of NFV Management and Orchestration (MANO) schedules computing and storage resources uniformly depending on the traffic demand. VIM slices the storage and computing resources and creates virtual machines (VMs) to carry functional components (flow splitting, mapping, to name a few).

We describe three main functional components in our scheme including flow splitting, mapping and flow recovery.

A. Flow Splitting

In this section, traffic is split into several traffic units by certain rules, where the level of splitting granularity can avoid network congestion by splitting into the smallest possible scale, i.e., a single packet. There are three types of traffic splitting, packet-level, flow-level and subflow-level, according to splitting granularity [3]. In this paper, the splitting granularity is subflow-level, so packet reordering is not taken into consideration.

Flow splitting is implemented in source node. All packets in a subflow are destined for the same path, but all packets heading for the same destination are carried in different paths. First, the SDN controller selects multiple paths and computes the weight of each path. Then, the SDN controller sends the path weight to source node. At last, flow splitting component identifies the first and last packet of each subflow.

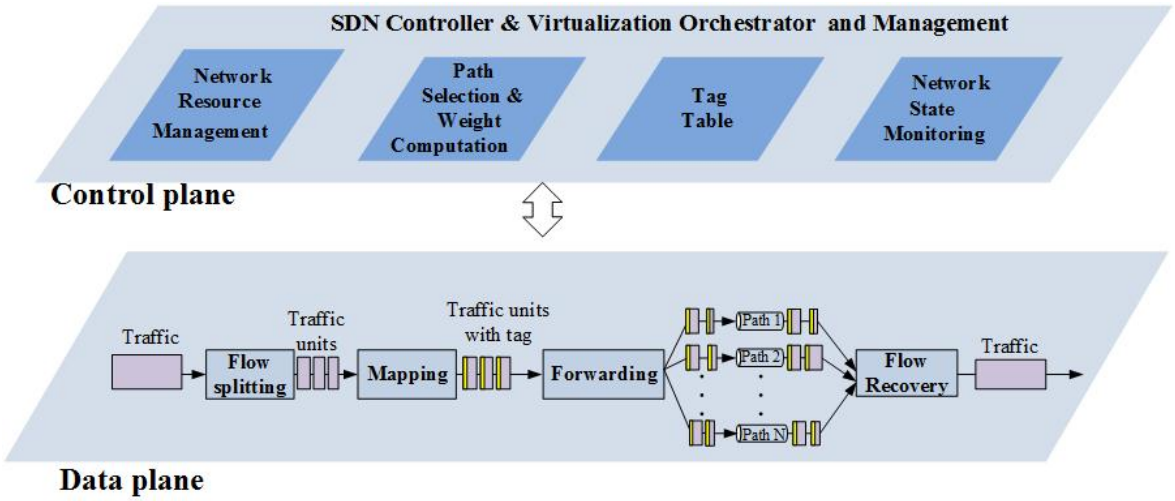

Fig. 1. The multipath network virtualizationimplementation scheme

B. *Mapping*

Mapping component aims at distinguishing subflows and forwarding based on tag matching. The original header of all packets in the subflow is replaced with a simplified tag embedded in OpenFlow header after the flow splitting is completed. For a flow, tag table will generate a set of tags based on subflows, and each tag corresponds to a path. Therefore, the number of tags is equal to the number of selected paths. Tag is valid only in the selected area by control plane. It is added in the source node and removed from the destination node.

The design of the tag is based on commonly observed phenomena of flows: (i)a flow takes multiple paths, (ii) a different flows have different priorities, and (iii) subflows need serial number. Therefore, for a subflow we just need to distinguish from which flow it comes and the flow priority. We also need to identify the path for a subflow and assign all the packets of subflow to this path. The OpenFlow 1.3.1 specification for SDN specifically allows information to be added to packets and additional headers by using the PUSH operation [17].

The tag format is shown in Fig. 2. It consists of Tag Identifier, Flow ID, Path ID, Packet Sequence Number and Priority. The tag identifier represents whether splitting granularity is packet-level or subflow-level. The flow ID takes up 16 bits and therefore could identify 2^{16} flows.

Tag Identifier	Flow ID	Path ID	Packet Sequence Number	Priority

Fig. 2.Tag format

C. *Flow Recovery*

The subflow is sorted by flow ID and path ID in flow recovery component after it reaches the destination node.

If the packet is lost, the destination node will send packet-in message to SDN controller. After completing the flow recovery, the tag will be replaced by the original header and then be sent to destination network. Meanwhile, flow recovery component notifies the tag table to remove the correspondence between tag and path.

III. DMMP ALGORITHM

Elephant flows account for less than 10% of all flows, but they carry more than 80% of the entire traffic volume [18].In this section, we present the Delay Minimization Multi-Path algorithm with network virtualization (DMMP) for elephant flow in detail.

A. *Network model& Solution formulation*

Assume the network is a graph $G=(V,E)$, where V is the set of nodes in the network and E is the set of links or edges. Each link $e=(v_i,v_j)\in E$， and a path p in G is defined as a list of consecutive nodes $(v_1,v_2,...,v_n)$, $\forall i$, $1\le i<n$ .The propagation delay and the minimum of the residual bandwidth of path p are denote by $d(p)$ and $w(p)$. We have

$$d(p)=\sum_{i=1}^{n-1}d(v_i,v_{i+1}) \tag{1}$$

$$w(p)=\min\{w(v_i,v_{i+1})\},1\le i<n \tag{2}$$

Define the set of multiple paths $P=\{p_1,p_2,...,p_K\}$, where K is the number of paths. The aggregate bandwidth of K path is,

$$W_K(P)=\sum_{m=1}^{K}w(p_m) \tag{3}$$

Further, define the required bandwidth of elephant flow as B_f .

Elephant flow is split into a group of subflow via K paths. If we consider that the destination acknowledges that the last subflow from the longest path arrives. The delay and average delay of the group j is defined as

$$D_j(P) = \max\{d(p_m)\} \tag{4}$$

$$\bar{t} = \frac{\sum_{m=1}^{k} d(p_m)}{k} \tag{5}$$

The multipath routing with virtualization problem can be presented as follows:

We can get the available bandwidth and propagation delay of each link via the network resource management module in control plane. Aggregate bandwidth should be larger than bandwidth required by elephant flow, which can ensure that packet loss rate and flow rate are not increased. We also choose the lowest delay of all feasible paths' set satisfying $W_K(P) \geq B_f$ as well as, $D_j(P)$ and $\bar{t}$ are minimized over all paths' set.

B. *Solution procedure*

In this section, we present the algorithm with network virtualization. We calculate the $W_K(P)$ of available K paths.

If $W_K(P) < B_f$, elephant flow will be sent into buffer and wait for $W_K(P) \geq B_f$.

If $W_K(P) \geq B_f$, we choose all multiple paths' sets satisfy in $W_k(P) \geq B_f$ and $W_{k-1}(P) < B_f$, where $2 \leq k < K$,

We get $D_j(P)$ and $\bar{t}$ from all elected paths' sets, and then the set satisfying $D_j(P)$ and $\bar{t}$ are minimized is been chosen.

The each path weight λ_m is shown as:

$$\lambda_m = \frac{w(p_m)}{W_k(P)} \tag{6}$$

IV. Experiments and Performance Evaluation

A. *Experimental platform*

We conduct multipath network virtualization experimental platform, which uses SDN and NFV technologies. OpenFlow switches supporting OpenFlow 1.3.1 protocol are chosen for forwarding. Nodes are connected by fiber links. Three industrial-level servers are deployed with Open platform for NFV(OPNFV) [19], and one server is deployed as control node which includes SDN controller and NFV orchestrator, and the other two servers are deployed as computing nodes in source and destination node. VMs are created on the computing nodes. One VM runs the flow splitting function, one VM runs the mapping function, and another one VM runs the flow recovery function. Two VMs run *Iperf* Client and Server to emulate elephant flows that have high-bandwidth demand. The remaining VMs which are added or removed based on the need of run OpenvSwitch (OVS) for forwarding. Besides, we deploy four severs at source and destination nodes. Two severs on each side are used to generate randomly varying background traffic. The network topology is shown is Fig.3. The configuration required for the experiment is shown in Table I. The DMMP algorithm is embedded in SDN controller, so SDN controller sends OpenFlow tables to source node, destination node and other nodes. SDN controller also sends weight of multiple paths to source node.

TABLE I. Hardware and Software Configuration

Device	Configuration
Server	Operating system: ubuntu 14.04 Intel Xeon processor @2.9GHz 64GB RAM Virtual Switch: OpenvSwitch SDN Controller: OpenDayLight (Lithium-SR3) NFV platform: OPNFV（Arno）
Switch	Switch model: Centec V350 OpenFlow specification 1.3.1 L2 to L4 complete matching fields 8*1Gbase-T 12*10Gbase-X

In the experiments, elephant flows have the same priority varying from 450 to 900Mbits in three paths and 2.1 to 3Gbits in ten paths. All the bottleneck bandwidth of paths is 500Mbps, background traffic varies randomly between 150-350Mbps. We separate background traffic and elephant flows by the action of Set-Queue in OpenFlow table. The delay is the maximum delay of selected paths. The splitting granularity is subflow-level, and paths are chosen to send flow using TCP protocol. The identifier of the designed tag takes up 1bit and settles binary 1 in tag identifier to represent as subflow-level, flow ID takes up 4bits, path ID takes up 2 bits, and priority takes up 2 bits. Through many times of experiments, we test the average delay, average throughput and average link utilization. Consequently, the splitting time and the delay of fiber are negligible compared to the propagation delay in the experiments.

The proposed scheme is compared with ECMP. *Wireshark* is used to capture all packets of elephant flow to calculate the delay, link utilization and throughput.

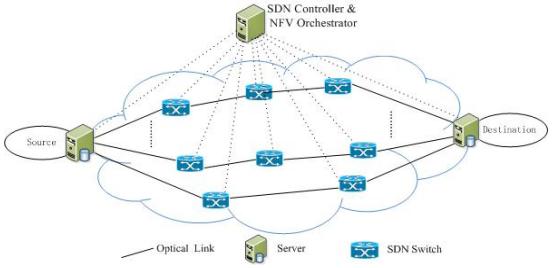

Fig.3 .Illustration of network topology in the experiment

Fig.4 and Fig. 7 show the average delay of elephant flows between DMMP and ECMP in the case of three and ten paths, respectively. From the two figures, we see that the trend of DMMP is approximately linear, because DMMP chooses multiple paths dynamically with the multipath network virtualization. But the delay of DMMP is no larger than 1 second under elephant flows is less than the available bandwidth, the experiment shows that the delay is more than 1 second. The main reason is that the available bandwidth of the link obtained by SDN controller is different from the actual situation. The trend of ECMP has dramatic fluctuation, especially in ten paths, because various background traffic result in the state of link difference and ECMP establishes multiple paths by the default way.

From Fig.5 and Fig.8, it is clear that DMMP has larger average throughput than ECMP. There is an increase of the average throughput as load increases in DMMP. While the average throughput of ECMP decreases as load increases in some points, because the throughput of ECMP is limited by the minimum available bandwidth in multiple paths. Multipath network virtualization provides the information of current network, and DMMP makes the most use of the available bandwidth, so the load imbalance is avoided and delay is stable.

From Fig.6 and Fig.9, we can see the average link utilization of DMMP is about 90% in two figures. The average link utilization of ECMP ranges from 65% to 82% in Fig.5 and 45% to 75% in Fig.8.The results show that the states of links become more complex as the number of links increases, and the disadvantages of ECMP are more prominent. Multipath network virtualization avoids the unreasonable use of available bandwidth.

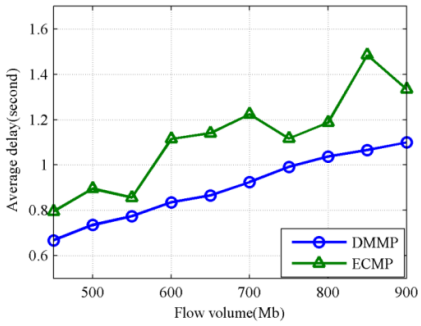

Fig.4.Performance comparison of average delay in three paths

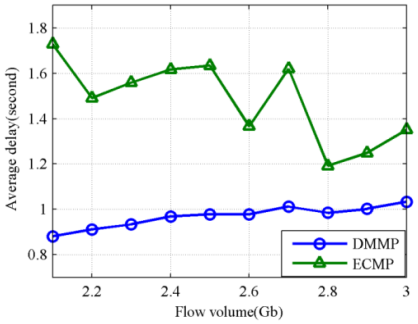

Fig. 7. Performance comparison of average delay in ten paths

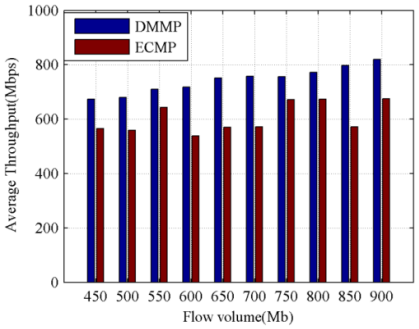

Fig.5.Performance comparison of average throughput in three paths

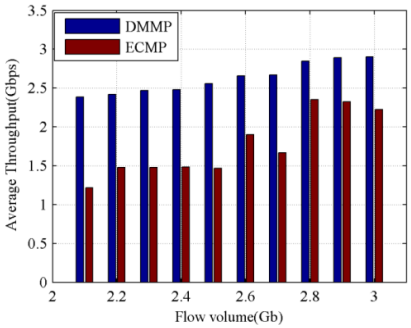

Fig.8.Performance comparison of average throughput in ten paths

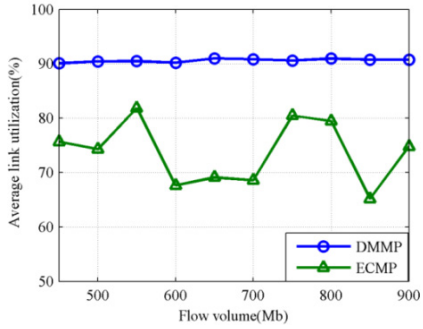

Fig.6.Performance comparison of average link utilization in three paths

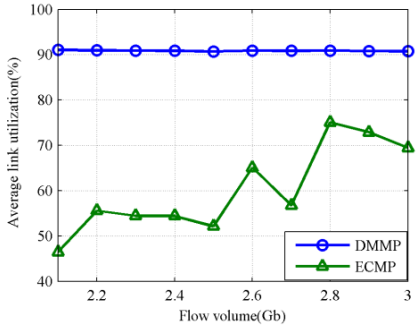

Fig.9.Performance comparison of average link utilization in ten paths

V. CONCLUSION

In this paper, we propose the multipath network virtualization implementation scheme with SDN and NFV. The scheme addresses the dilemma of multipath implementation. In this scheme, SDN controller schedules the network resources to select multiple paths and compute the weight of each path according to the requirement of flow and dynamic network state. NFV orchestrator manages computing and storage resources to split flow, add tag and recover flow. We also propose a multipath algorithm with our scheme for elephant flow. Experiments on multipath network virtualization platform are conducted. The results show our proposed algorithm applied on multipath network virtualization experimental platform has superior performance than ECMP. In the future, we will optimize the scheme and set up multipath mathematical modeling with network virtualization.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (Grant No. 61471053), the 111 project (Grant No. B17007) and Director Funds of Beijing Key Laboratory of Network System Architecture and Convergence (Grant No. 2017BKL-NSAC-ZJ-02).

REFERENCES

[1] J. R. Iyengar, P. D. Amer, and R. Stewart, "Concurrent multipath transfer using SCTP multihoming over independent end-to-end paths,"IEEE/ACM Trans. Netw., vol. 14, no. 5, pp. 951–964, Oct. 2006.

[2] M. Handley, A. Ford, C. Raiciu, and O. Bonaventure, "TCP extensions for multipath operation with multiple addresses," IETF, Fremont, CA, USA, RFC 6824, 2013.

[3] SumetPrabhavat,HirokiNishiyama,NirwanAnsari,andNeiKato, "On Load Distribution over Multipath Networks,"IEEE Communications Surveys & Tutorials, vol. 14, no. 3, pp.2157-2175, 2012.

[4] J. Moy, "OSPF version 2," RFC 2328, Apr. 1998.

[5] R. CalIon, "Use of OSI IS-IS for routing in TCP/IP and dual environments," RFC 1195, Dec. 1990.

[6] S. Kandula, D. Katabi, S. Sinha, and A. Berger, "Dynamic loadbalancing without packet reordering," ACM SIGCOMM Computer Communication Review, vol. 37, no. 2, pp. 53–62, Apr. 2007.

[7] W. Shi, M. H. MacGregor, and P. Gburzynski, "Load balancing forparallel forwarding," IEEE/ACM Trans. Netw., vol. 13, no. 4, pp. 790–801, Aug. 2005.

[8] A. Zinin,"Cisco IP routing: packet forwarding and intra-domain routingprotocols,"Addison-Wesley, 2002.

[9] Shan He, GuochuShou, Yihong Hu, ZhigangGuo, " Performance of Multipath in Fiber-Wireless(FiWi)Access Network with Network Virtualization,"Military Communications Conference, MILCOM 2013 - 2013 IEEE, pp.18-20, Nov. 2013.

[10] AnjingWang,MohanIyer, RudraDutta, George N. Rouskas, and Ilia Baldine, "Network Virtualization: Technologies,Perspectives, and Frontiers,"Journal of Lightwave Technology, vol. 31, no. 4, pp.523-537, Feb.2013.

[11] M. Maier and B. P. Rimal, "Invited paper: The Audacity ofFiber-Wireless (FiWi) networks: Revisited for Clouds and Cloudlets," China Commun., vol. 12, no. 8, pp. 33–45, Aug. 2015.

[12] ETSI, "Network Functions Virtualisation: Architectural framework, standard no. GS NFV 002 v1.2.1," Dec. 2014.

[13] Ericsson, "TelefonicaandEricssonpartnertovirtualizenetworks,pressrelease ,"Websitehttp://www.ericsson.com/news/1763979. Feb. 2014.

[14] Thi-Thuy-LienNguyen,Tuan-MinhPham,HuynhThiThanhBinh,"Adaptive Multipath Routing for Network Functions Virtualization,"SoICT '16 Proceedings of the Seventh Symposium on Information and Communication Technology, pp. 222-228.

[15] S.Q.Zhang,Q.Zhang,H.Bannazadeh,andA.Leon-Garcia, "Routing algorithms for network function virtualization enabled multicast topology on SDN," IEEE Transactions on Network and Service Management,vol.12,no.4,pp.580-594,Dec.2015.

[16] Tuan-Minh Pham,LinhManhPham, "Load Balancing using Multipath Routing In Network Functions Virtualization,"The 2016 IEEE RlVF International Conference on Computing & Communication Technologies,Research,Innovation,and Vision for the Future, Nov. 2016.

[17] OpenFlowSwitchSpecification,Version1.3.1.https://www.opennetworking.org/images/stories/downloads/sdnresources/onf-specifications/openflow/openflow-spec-v1.3.1.pdf.

[18] Curtis R A,Mogul C J,TourrihesJ,etal."DevoFlow:Scaling Flow Management for High-Performance Networks,"ACM SIGCOM Computer Communication Review, vol.41, no.4, pp.254-265, 2011.

[19] https://www.opnfv.org

[20] Jing Liu, Jie Li, etal. "SDN Based Load Balancing Mechanism for Elephant Flow in Data Center Networks,"2014 International Symposium on Wireless Personal Multimedia Communications (WPMC2014),Sep.2014.

[21] Junlan Zhou,Malveeka Tewari,etal. "WCMP:weighted cost multipathing for improved fairness in data centers,"Proceedings of the Ninth European Conference on Computer Systems,Apr. 2014.

[22] RodolfoAlvizu,GuidoMaier,MassimoTornatore,etal."Differential delay constrained multipath routing for SDN and optical networks,"Electronic Notes in Discrete Mathematics,vol.52,pp.277-284,2016.